Pre-trainning for Recommender System:A Survey

QinHsiu

***Abstract*—The purpose of recommendation systems is to provide better item recommendation to users, but they often face the problem of lack of data (cold start) in practical application scenarios. Recently, pre-training models have achieved good results in different domains and tasks. Introducing pre-training mechanism in recommendation models not only alleviates the problem of lack of data in recommender systems, but also improves the model results. This paper provides an overview of results and research on the use of pre-training for recommendation, and discusses promising future research directions for using pre-training mechanisms for recommender systems.**

***Index Terms*—Recommender System, Pre-training, Cold Start, Fine Tune, Self-supervised**

I. INTRODUCTION

WITH the rapid development of the Web, users are faced with the problem of information overload, and the large amount of online product information makes it difficult for users to make effective decisions. The emergence of recommendation systems can alleviate this problem. Rcommender system is a filtering system that presents personalized information to users, which not only improves the user experience, but also increases business benefits as a result.

Recommender systems usually face the problem of sparse data in real life. For example, recommender systems suffer from cold-start problems when faced with new users or new products [1]. Pre-trained models are usually trained on unlabeled data to learn universal linguistic representations, and then knowledge transfer is obtained in downstream tasks, which can alleviate the data sparsity problem to some extent.

In the context of recommender systems, models that use pre-training mechanisms to improve recommendation accuracy can be divided into two categories (as shown in Figure 1): feature-based models and fine-tuning-based models. Feature-based models use pre-training to obtain features from edge information. Fine-tuning-based models use the user-item interaction records to pre-train a deep migratable neural network model, and then use different downstream tasks to fine-tune the pre-trained model parameters. In general, the benefits of using pre-training in recommender systems can be summarized as follows:

(1) Pre-training tasks can better explore the information of user-item interactions and thus better capture user interests.

(2) Pre-training can help integrate knowledge from different tasks to obtain a more general representation of users and products, which can be further adapted to different recommendation scenarios.

The main contributions of this paper are as follows:

(1) A systematic classification and summary of models of pre-training methods for recommender systems is presented.

(2) A variety of promising research directions for future research are discussed.

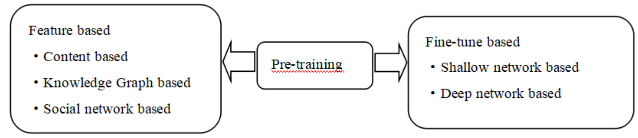

Fig. 1

The remainder of this review is described as follows: an overview of existing models of pretraining for recommendation is presented in Section 2, and future research directions are discussed and summarized in Section 3.

II. RESEARCH STATUS

A. Feature based model

The main feature of feature-based models is the introduction of auxiliary information (e.g., item attributes, knowledge graphs, or social networks) that is used to enrich the representation of user and item information using pre-trained models that learn directly from the auxiliary information. Unlike collaborative filtering, which learns representations in the user-item interaction records, feature-based models focus on using pre-trained models to learn from the large number of features that already exist and can be used, and integrate these features into downstream recommendation tasks. By combining rich auxiliary information with interaction information about user goods, feature-based models can alleviate the potential data sparsity problem. Feature-based recommendation models preprocess auxiliary information through diverse pretrained models to obtain embedded representations of users and products, and then integrate these embedded representations into downstream recommendation tasks to enrich the representations of users and products for better personalized recommendations.

Because different types of auxiliary information require different pre-training models for training. Based on the types of external resource information, feature-based pre-training models can be classified into the following categories, including content-based recommendation, knowledge graph-based recommendation, and social recommendation models. The details of how the pre-training mechanism is used in these three types of recommendation models are described next.

1) Content based model: Content-based recommendations assume that users tend to purchase items similar to those they have already purchased. Therefore, it is important for content-based recommendation systems to encode item information into a more symbolic, low-dimensional representation.

He and McAuley [2] were the first to incorporate visual features into recommendation tasks, using pre-trained models to extract visual features from product images for downstream recommendation tasks. song et al [3] focused on combining long- and short-term user interests for improving recommendations. zhang et al [4] were the first to propose a framework that is easily scalable to new information sources to avoid in-practice retraining the model.Zheng et al [5] propose to learn both product attribute representation and user behavior representation from review text.Brochier [6] with Cenikj and Gievska [7] simply use word embeddings to represent the whole text, thus deepening the focus on text attributes with some results.

2) Knowledge Graph based model: Knowledge graph-based recommendation introduces knowledge graphs as supplementary information to better capture the characteristics of users and items. A knowledge graph is a structured graph containing a large amount of factual information and connection information between users, commodities and other connected entities. A large amount of auxiliary information, such as user profiles, commodity attributes, or cross-domain commodity connections, can be integrated into a knowledge graph. Thus, knowledge graphs can help recommendation systems by capturing essential knowledge and providing interpretability of recommendation results.

So far, a large number of knowledge graphs have been used in different works. To develop knowledge graphs, a series of knowledge graph-based approaches attempt to encode knowledge graphs as low-dimensional pre-trained embeddings using knowledge graph embedding methods. For example, Bordes et al [8], Wang et al [9] and Lin et al [10] have used this approach. Knowledge graphs capture user and product features by aggregating embedding information from interaction data. For example, some works [11,12] build knowledge graphs by commodities and attributes associated with them. Some other works [13,14] add user information to build user-commodity graph information containing commodity attributes, user behavior information (e.g., purchases, clicks, etc.), and user portraits. Using information-rich heterogeneous user-goods graphs, the potential relationships between users and goods can be directly modeled.Wang et al [15] proposed a model for knowledge graph attention networks that explicitly models higher-order connectivity in knowledge graphs in an end-to-end manner, and obtained good experimental results by modeling the display of higher-order relationships.The experimental results of Qin et al [16] showed that that knowledge graphs are effective sources of information and can significantly improve recommendation effectiveness.

Chen [17] et al. first proposed a novel joint nonsampling learning model for enhanced knowledge graph recommendation.Chen [17] et al. first designed a new efficient negative sampling optimization algorithm for knowledge graph embedding learning. The subgraphs are encoded by an attentional neural network to better characterize users' preferences for items. By cleverly designing a memory strategy and a joint learning framework, the model not only models the fine-grained connections between users, items, and entities, but also efficiently learns model parameters from the entire training data (including all unobserved data).Wang [18] et al. first proposed a new negative sampling strategy by incorporating knowledge graphs into negative sampling, which acts as a reinforced learning agent to explore high-quality negative information.Wang [19] et al. designed a distinguishable sampling strategy that enables the selection of relevant goods to be jointly optimized with the training process of the model. The strategy learns the relevance distribution of connected items from the knowledge graph and samples suitable items for recommendation according to that distribution.Zhao et al [20] proposed a new model that takes co-occurrence information from the original data into account for enriching the representation of goods and users.

3) Social recommendation: Social recommendation is a recommendation method that uses social connections as an additional complementary input. Unlike knowledge graphs that integrate diverse information about users and goods, social relationship graphs focus on modeling relationships among users. Homogeneity theory suggests that users' preferences are similar to or influenced by their social friends... Similar to knowledge graph-based recommendations, many social recommendation models seek to integrate pre-trained social network embeddings, which indicate the extent to which users are influenced by their friends.

Guo [21] proposed a model consisting mainly of an embedding model and a filtering model.Zhang et al [22] proposed a new framework that combines network embedding with matrix decomposition.Chen et al [23] proposed a new model that considers both high and low similarity between users.

B. *Fine-tune based model*

The fine-tuned models used for recommendations are first pre-trained on large-scale pre-training data. The pre-trained models are migrated to downstream tasks, using a small amount of data to fine-tune the existing model parameters. This paradigm of fine-tuning has demonstrated its effectiveness in other domains. Examples include natural language processing [24,25] and computer vision [26]. Based on the structure of the models, the existing models can be divided into two categories: shallow neural networks and deep residual neural networks. The already existing deep neural neural networks can be further classified into recommended models based on Bidirectional Encoder Representations from Transformers (BERT) and parameter efficient pre-trained convolutional neural networks.

1) Shallow Neural network: Initial work attempted to capture knowledge transfer through shallow neural networks as the underlying model, e.g., shallow multilayer perceptron, recurrent neural networks. Hu et al [27] tried to improve recommendation by sharing knowledge across domains. They conducted a baseline experiment on a multilayer perceptron where the multilayer perceptron was embedded with users and goods in the original domain. Users are embedded in the target domain. The experimental results show that this simple approach cannot obtain significant improvements in the recommendation domain. The final results of the experiments show that the model results and the pre-training task need to be designed in such detail that effective knowledge transfer can be obtained in the fine-tuned model.

Ni et al [28] proposed the Deep User Perception Network (DUPN) model, which is capable of learning universal user representations through multiple recommendation tasks.DUPK takes interaction sequences as input and later uses long and short-term memory and attention layers to obtain user representations.DUPN is pre-trained with multi-task objects for including click-through rate prediction, shopping performance prediction, and price prediction. Experimental results show that DUPN not only helps from pre-training, but also converges faster and has better results in relevant new tasks. Although the user representation learned by DUPN is useful, DUPN needs many additional information resources such as user portraits to implement different pre-training tasks.

Zhou et al [29] proposed a new sequential recommendation model S3Rec, which uses four prediction tasks to capture the relationships between sequences during pre-training, and the pre-trained model parameters are directly used for fine-tuning in downstream tasks.Liu et al [30] proposed to use pre-training to reverse predict data for the purpose of expanding data and learning commodity representations, and then the resulting enhanced data and models are then used for downstream recommendation tasks.

Cheng et al [31] first used contrast pre-training to obtain user representations for downstream tasks, proposing a new contrast learning pre-training framework that uses contrast learning to model sequential user behaviors, treating each user behavior as a whole and transforming the original user behavior by means of data augmentation to construct self-supervised signals and thus learn better user representations.

2) BERT based model: To capture dynamic user preferences, many studies have attempted to explore temporal sequence-based user sequences, mainly in the form of session-based recommendations. Similar to natural language processing, which targets word sequences, conversation-based recommendations take sequence information into account when analyzing sequences of goods. Motivated by the rapid development of pre-trained language models in the field of natural language processing [24, 25], much of the work is turning to the use of pre-trained models to capture information in sequences of user behavior, especially BERT-based models. In this section, pre-trained BERT models for recommendation are presented, including the most widely used Masked Item Predict (MIP), the structure of BERT, and BERT models for recommendation.

MIP is widely used in many recommender systems as in the masked language model of natural language processing. In the masked commodity prediction task, given an interaction sequence, some of the commodities are masked randomly. A pre-trained model is used to reconstruct these masked goods. As an example, we label the time-based order information of user u Su=v1,v2,...,vn, where vi denotes the goods that user u interacts with at position i of time step. For the pre-training task, some input goods present in the interaction sequence are randomly masked and replaced with [MASK]. This model is used to predict these masked commodities. The goal of this task is to minimize the prediction error. Unlike the task of left-to-right next commodity prediction used in many conversational recommender systems, masked commodity prediction allows the model to learn a representation of the sequence of user behavior from the entire context.Sun et al [32] show that the masked commodity prediction task can overcome the limitation that strict sequence order in user behavior is not always effective. Thus, pre-trained models using a masked commodity prediction task can yield better results.

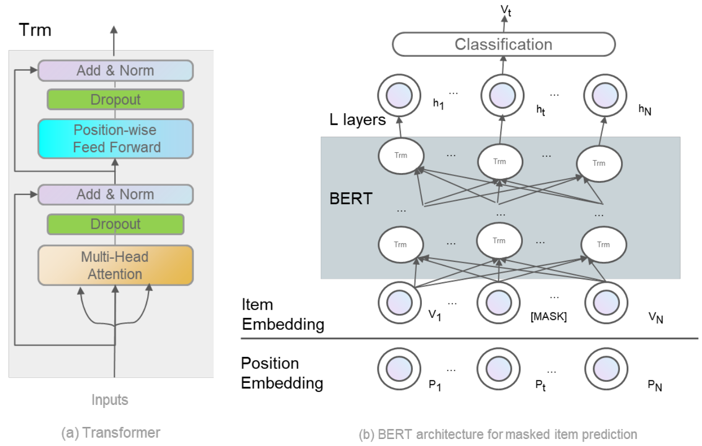

Fig. 2: Fig(a) is a transformer structure and Fig(b) is a model of BERT for MIP

The pre-trained BERT model labeled for the masked commodity prediction task. As shown in Fig. 2 (b), BERT is based on a multilayer, bidirectional migration proposed by Vaswani et al [33], where the migration model is shown in Fig. 2 (a) and includes a multi-headed attention sublayer with a feedforward neural network.

Sun et al [32] trained a two-layer BERT to accomplish masked commodity prediction and obtained best-of-the-time (State of the Art, SOTA) performance in the next sequence-based commodity recommendation task. They observed that the structure of BERT and the task of masking commodity prediction could significantly improve performance, and that multiple migration layers could further enhance performance on large datasets, providing foundational support for later models. In the same year Chen et al [34] used BERT for conversational recommendation and also achieved good results.Qiu et al [35] proposed the model U-BERT, which introduced the task of masked view prediction as well as user view hierarchy prediction during pre-training to enrich the user's representation by introducing review information.The model was pre-trained and then fine-tuned by setting different downstream tasks to achieve good results. A more applicable model, UPRec, was proposed by Xiao et al [36], which used three subtasks to complete the pre-training, namely, masked goods prediction, user portrait prediction, and social relationship prediction. The latter two of these three tasks are collectively referred to as the user awareness task in the text. By introducing the latter two tasks, the model enriches the user representation and obtains SOTA results as a result.

3) Parameter Efficient model: The pre-training mechanism enables the model to capture the user's representation from the user's behavioral history through a self-supervised learning approach. Experimental results show that this can be used for recommendations to obtain significant improvements. However, fine-tuning the model separately for different tasks is computationally and storage expensive

To address this problem, Yuan et al [36] proposed the parameter-efficient transfer learning architecture (PeterRec), which uses grafted neural networks called model patches during fine-tuning, and by inserting the model patches into the pre-trained model, the fine-tuned The model is able to

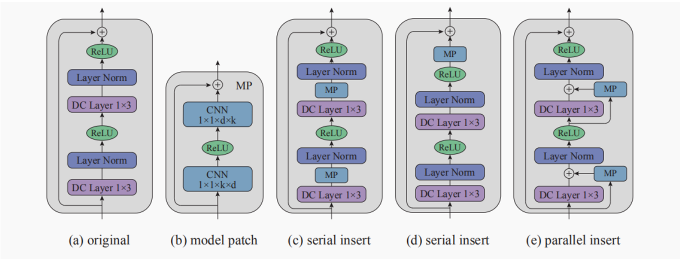

Fig. 3: Model patch (MP) and insertion methods. (a) is the original pre-trained residual block; (b) (c) (d) (e) are the fine-tuned residual blocks with inserted MPs; and (f) is the MP block. + is the addition operation. 1×3 is the kernel size of dilated convolutional layer.

guarantee that all pre-trained parameters remain unchanged. PeterRec is an expansive convolutional layer and uses a residual connection between every two expansive convolutional layers, as shown in Image 2. Similar to other pre-training models, PeterRec uses masked commodity prediction for pre-training. During fine-tuning, a model patch (a two-layer residual convolutional network), is inserted around the original dilated convolutional layers. The parameters of the pre-training are shared across tasks. Only the parameters of the model patch are fine-tuned. In order to speed up the fine-tuning and minimize the number of parameters, the model patch block is designed as a bottleneck structure. In particular, the first convolutional layer projects the k-dimensional channels onto the d(d is much smaller than k)-dimensional vectors and the second layer maps them to the original dimensions. In this way, the inserted parameters are reduced by 10% compared to those of the original pre-trained model. Experimental results show that the pre-trained PeterRec is effective for a variety of downstream tasks, including user portrait prediction and up to k recommendations. Moreover, it is able to obtain significant results when cold-start problems are encountered, which demonstrates the effectiveness of pre-trained models for knowledge migration in recommender systems.

III. Conclusion and Future Prospect

A. Challenge

Although pre-trained models have demonstrated their effectiveness in the field of recommender systems, challenges persist. This section provides an introduction to five challenges and future directions for recommender systems. The first three of these subsections discuss how to better use pretrained models in diverse recommendation scenarios, while the next two subsections focus on how to improve the use of pretraining mechanisms to better serve recommender systems.

1) Cold Start: Collaborative filtering recommender systems rely mainly on historical user interaction data and suffer from the cold start problem. To alleviate this problem, some research workers [37-39] proposed using auxiliary information, such as user portraits and product attributes, to enrich the representation of users and products. In addition to this, some use more efficient learning mechanisms to alleviate the heavy reliance on data, such as small sample learning [39,40]. In recommendation, pre-trained models are able to learn migratable representations of shared information from large-scale other domains with sparse target domains, which can be used for cold-start problems. For example, if a user is cold-started in the target domain, it is useful to pre-train a representation of the user in the general domain. If a good is cold-started, the representation of the good can be inferred from a pre-trained representation of external information.Peterrec [36] explains well that the use of pre-trained models can alleviate the problem of cold-starting of users.

2) Knowledge enhanced pre-training: Knowledge graphs can provide rich domain knowledge, world knowledge and common sense knowledge for recommendation. Thus, by introducing knowledge graphs in recommendations, the connection between user preferences and goods can be captured more accurately. Some work directly considers putting external structural knowledge into pre-trained models for recommendation. In fact, many knowledge-enhanced pre-trained language models [42,43] have shown that mixing structured knowledge into pre-trained models can significantly improve the performance of the original models. The knowledge information can help the model to better learn the representation of users and products and, by doing so, improve the performance of recommendations.

3) Social relation enhanced pre-training: Social connections provide a possible perspective for personalized recommendations. Users who are connected are more likely to share similar preferences. Pre-trained models are skilled at capturing users' interests from their history of interactions. Thus, social relationships among users can be viewed as meta-relationships between users and goods. Meta-relationships between interaction sequences, i.e., interaction sequences of closely connected users are encouraged to share similar representations. That is, interaction sequences of closely connected users are encouraged to share similar representations. Based on this, sequence-level pre-training tasks can be proposed to help the model produce more expressive user/item representations.

Another possible direction is to use social relationship-enhanced pre-trained models to solve the user cold-start problem. Social relationships can provide clues to users' interests. However, it is still a challenge to make full use of the rich information of neighboring users during the pre-training process.

4) pre-training task: Currently, most of the deep fine-tuning approaches rely on MIP tasks to pre-train models. Instead, these efforts focus on extracting user interests from the user's historical continuous records. However, limited by the computational power and memory of GPUs, only the most recent interaction records, which represent the most recent user preferences, can be utilized by recommendation models. Moreover, MIP can only utilize continuous data, while there is usually rich heterogeneous information in real-world scenarios. Therefore, designing new self-supervised pre-training tasks is important to take full advantage of the large-scale heterogeneous information.

5) model structure: Pre-trained models can be used for a variety of downstream tasks, but their high computational complexity makes them difficult to apply to real-world situations. In addition, fine-tuning each downstream task individually is quite time-consuming and memory-consuming. How to achieve fast and efficient knowledge transfer is still an urgent problem to be solved.

B. Conclusion

In this paper, we study pre-training models for recommendation and summarize the work in this area. We provide a more comprehensive overview of two types of pre-training models for recommendation, feature-based models and fine-tuning-based models. We then discuss the challenges and promising future directions of pretraining for recommendation, which we hope will contribute to the development of the field.

References

[1] J. Gope and S. K. Jain, "A survey on solving cold start problem in recommender systems," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*. Greater Noida: IEEE, May 2017, pp. 133–138.

[2] R. He and J. McAuley, "Vbpr: visual bayesian personalized ranking from implicit feedback," in *Association for the Advancement of Artificial Intelligence(AAAI)*, 2014, pp. 144–150.

[3] H. X. Song Y, Elkahky A M, "Multi-rate deep learning for temporal recommendation," in *39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 2016, pp. 909–912.

[4] C. X. e. a. Zhang Y, Ai Q, "Joint representation learning for top-n recommendation with heterogeneous information sources," in *2017 ACM on Conference on Information and Knowledge Management.*, 2017, pp. 1449–1458.

[5] Y. P. S. Zheng L, Noroozi V, "Joint deep modeling of users and items using reviews for recommendation," in *the tenth ACM international conference on web search and data mining.*, 2017, pp. 425–434.

[6] B. R., "Representation learning for recommender systems with application to the scientific literature," in *Companion Proceedings of The 2019 World Wide Web Conference.*, 2019, pp. 12–16.

[7] G. S. Cenikj G, "Boosting recommender systems with advanced embedding models," in *Companion Proceedings of the Web Conference 2020*, 2020, pp. 385–389.

[8] G.-D. A. e. a. Bordes A, Usunier N, "Translating embeddings for modeling multi-relational data," in *the Advances in Neural Information Processing Systems.*, 2013, pp. 2787–2795.

[9] F. J. e. a. Wang Z, Zhang J, "Knowledge graph embedding by translating on hyperplanes," in *the 28th AAAI Conference on Artificial Intelligence.*, 2014, pp. 1112–1119.

[10] S. M. e. a.]Lin Y, Liu Z, "Learning entity and relation embeddings for knowledge graph completion," in *the 29th AAAI Conference on Artificial Intelligence.*, 2015, pp. 2181–2187.

[11] D. H. e. a. Huang J, Zhao W X, "Improving sequential recommendation with knowledge-enhanced memory networks," in *the 41st International ACM SIGIR Conference on Research Development in Information Retrieval.*, 2018, pp. 505–514.

[12] H. M. e. a. Wang H, Zhang F, "Shine: Signed heterogeneous information network embedding for sentiment link prediction," in *the Eleventh ACM International Conference on Web Search and Data Mining.*, 2018, pp. 592–600.

[13] R. O. e. a. Dadoun A, Troncy R, "Location embeddings for next trip recommendation," in *Companion Proceedings of The 2019 World Wide Web Conference.*, 2019, pp. 896–903.

[14] H. X. e. a. Cao Y, Wang X, "Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences," in *The world wide web conference.*, 2019, pp. 151–161.

[15] C. Y. e. a. Wang X, He X, "Kgat: Knowledge graph attention network for recommendation," in *the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining.*, 2019, pp. 950–958.

[16] Q. C. e. a. Guo Q, Zhuang F, "A survey on knowledge graph-based recommender systems," in *IEEE Transactions on Knowledge Data Engineering*, 2020 (01), pp. 1–1.

[17] M. W. e. a. Chen C, Zhang M, "Jointly non-sampling learning for knowledge graph enhanced recommendation," in *the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.*, 2020, pp. 189–198.

[18] H. X. e. a. Wang X, Xu Y, "Reinforced negative sampling over knowledge graph for recommendation," in *The Web Conference 2020.*, 2020, pp. 99–109.

[19] Wang Y, Liu Z, "Dskreg: differentiable sampling on knowledge graph for recommendation with relational gnn," in *the 30th ACM International Conference on Information Knowledge Management.*, 2021, pp. 3513–3517.

[20] Z. L. e. a. Zhao X, Cheng Z, "Ugrec: Modeling directed and undirected relations for recommendation," in *arXiv preprint arXiv:2105.04183*, 2021.

[21] C. Z. e. a. Wen Y, Guo L, "Network embedding based recommendation method in social networks," in *Companion Proceedings of the The Web Conference 2018.*, 2018, pp. 11–12.

[22] S. C. e. a. Zhang M, Hu B, "Matrix factorization meets social network embedding for rating prediction," in *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data.*, Springer, Cham, 2018, pp. 121–129.

[23] H. J. e. a. Chen J, Chen W, "Co-purchaser recommendation based on network embedding," in *International Conference on Web Information Systems Engineering.*, Springer, Cham, 2020, pp. 197–211.

[24] L. K. e. a. Devlin J, Chang M W, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, 2019, pp. 4171–4186.

[25] L. Y. e. a. Joshi M, Chen D, "Spanbert: Improving pre-training by representing and predicting spans," in *Transactions of the Association for Computational Linguistics*, 2020, pp. 64–67.

[26] C. Y. e. a. Su W, Zhu X, "Vl-bert: Pre-training of generic visual-linguistic representations," in *International Conference on Learning Representations*, 2019.

[27] Y. Q. C. Hu G, Zhang Y, "Conet: Collaborative cross networks for cross-domain recommendation," in *the 27th ACM international conference on information and knowledge management.*, 2018, pp. 667–676.

[28] L. S. e. a. Ni Y, Ou D, "Perceive your users in depth: Learning universal user representations from multiple e-commerce tasks," in *the 24th ACM SIGKDD International Conference on Knowledge Discovery Data Mining.*, 2018, pp. 596–605.

[29] Z. W. X. e. a. Zhou K, Wang H, "S^3-rec: Self-supervised learning for sequential recommendation with mutual information maximization," in *the 29th ACM International Conference on Information Knowledge Management*, 2020, pp. 1893–1902.

[30] W. Y. e. a. Liu Z, Fan Z, "Augmenting sequential recommendation with pseudo-prior items via reversely pre-training transformer," in *arXiv preprint arXiv:2105.00522*, 2021.

[31] L. Q. e. a. Cheng M, Yuan F, "Learning transferable user representations with sequential behaviors via contrastive pre-training."

[32] W. J. e. a. Sun F, Liu J, "Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer," in *the 28th ACM international conference on information and knowledge management.*, 2019, pp. 1441–1450.

[33] P. N. e. a. Vaswani A, Shazeer N, "Attention is all you need," in *Advances in neural information processing systems.*, 2017, pp. 5998–6008.

[34] L. C. e. a. Chen X, Liu D, "Bert4sessrec: Content-based video relevance prediction with bidirectional encoder representations from transformer," in *the 27th ACM International Conference on Multimedia.*, 2019, pp. 2597–2601.

[35] G. J. e. a. Qiu Z, Wu X, "U-bert: Pre-training user representations for improved recommendation," in *the AAAI Conference on Artificial Intelligence.*, 2021, pp. 4320–4327.

[36] Y. Y. e. a. Xiao C, Xie R, "Uprec: User-aware pre-training for recommender systems," in *arXiv e-prints arXiv: 2102.10989*, 2021.

[37] L. C. e. a. Ma H, Zhou D, "Recommender systems with social regularization," in *the fourth ACM international conference on Web search and data mining.*, 2011, pp. 287–296.

[38] O. I. Manotumruksa J, Macdonald C, "Regularising factorised models for venue recommendation using friends and their comments," in *the 25th ACM International on Conference on Information and Knowledge Management.*, 2016, pp. 1981–1984.

[39] L. J. e. a. Yu J, Gao M, "Adaptive implicit friends identification over heterogeneous network for social recommendation," in *the 27th ACM international conference on information and knowledge management.*, 2018, pp. 357–366.

[40] L. K. e. a. Li J, Jing M, "From zero-shot learning to cold-start recommendation," in *the AAAI Conference on Artificial Intelligence*, 2019, pp. 4189–4196.

[41] J. S. e. a. Lee H, Im J, "Melu: Meta-learned user preference estimator for cold-start recommendation," in *the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining.*, 2019, pp. 1073–1082.
[42] L. Z. e. a. Zhang Z, Han X, "Ernie: Enhanced language representation with informative entities," in *the 57th Annual Meeting of the Association for Computational Linguistics.*, 2019, pp. 1441–1451.
[43] Z. Z. e. a. Liu W, Zhou P, "K-bert: Enabling language representation with knowledge graph," in *the AAAI Conference on Artificial Intelligence.*, 2020, pp. 2901–2908.